Interactive Networks for Digital Cultural Heritage Collections: Scoping the Future of histoGraph

Marten Düring, Lars Wieneke
University of Luxembourg, Luxembourg
marten.duering@uni.lu, lars.wieneke@uni.lu

Abstract

Network visualizations are powerful albeit suggestive means to represent and explore complex relations between entities (people, words, places…) and have become increasingly popular with cultural heritage institutions. In this article we briefly introduce the functionality of histoGraph (Wieneke et al. 2014), a technical demonstrator for the network-based analysis of photographs, survey current applications of network visualizations in the cultural heritage domain, and outline histoGraph's future development as an open source tool for the visualization of cultural heritage data.

1 histoGraph

histoGraph has been developed in the context of the FP7-funded project CUbRIK (*CUbRIK Project* 2014) which focused on advanced multimedia search technologies. The current version creates a social network of persons who appear in photos related to the history of European integration and automatically enriches the network with relevant sources based on keyword queries in full text. A demo is available online (*histoGraph*). To create the network, the faces of the persons in the historical photos need to be identified which provides a very challenging task for machines alone. To improve the quality of the output, the task is therefore shared between humans and machines and the input from one improves the performance of the other.

histoGraph introduces an effective interface to access collections of historical sources and to discover links among and entities within them. The integration of human expertise and machine computation opens up new possibilities to create a new type of applications. So far however, this potential remains largely untapped because of the significant requirements for such projects: The implementation and integration of advanced algorithms, for example for the identification of faces, requires specialized know-how and users from the humanities are challenged with defining unprecedented tasks for methods which haven't even emerged yet. histoGraph combines new approaches to engage the public to commit to humanities research, to facilitate exchange between users and to help us reach our audiences.

CUbRIK integrated research in computer science, the design of human-computation tasks, data visualization, social engineering and the humanities. Multimedia search is still a challenge and the CUbRIK project approached it from many different angles: Human-Machine-Interaction, face identification in photos and videos, copyright issues, gamification, content exploration and user communication. Alongside an app for exploring and searching fashion, histoGraph is one of two demos which implement the different modules developed by CUbRIK.

histoGraph is based on a curated collection of more than 3000 images which represent the main events and actors in the history of European integration. This image collection is hosted by the CVCE. To prepare the photos for the network, we use an image indexation pipeline which detects the location of individual faces in the photographs. A crowd of "click-workers" with no specific training double-checks whether the algorithms detected faces correctly or whether it missed some. In the next step, an automatic face recognition process is triggered that associates each of

the now verified faces with a list of ten possible identities. This list of candidates is then disseminated for example through Twitter to a crowd of experts who vote for and comment on their preferred identity. The image metadata, for example the names of persons, the time or the place where an image was taken as well as contextual information about associated historical events can be reviewed by expert users and delegated to a crowd of specialists on the history of European Integration for review.

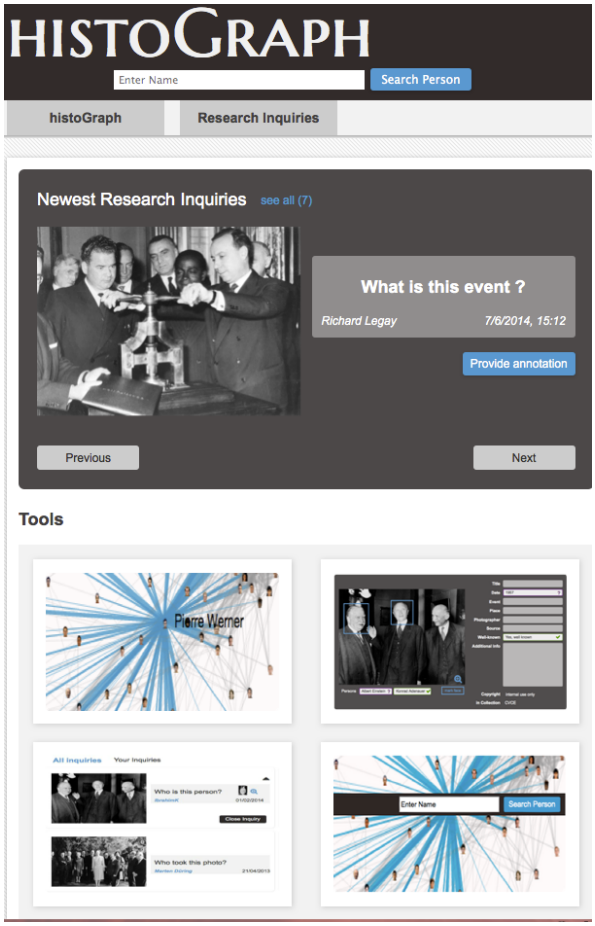

Figure 1: Screenshot of the histoGraph demonstrator.

Based on the co-occurrence of persons in images, a social network is calculated which links individual persons with each other (see Figure 2.). Connections gain in strength the more often persons appear together in an image. Users can interact with histoGraph in different ways, e.g. a click on a node leads to an ego-network of the selected person and a click on an tie displays documents which mention both actors (see Figure 3.).

This feature is powerful since it guides users back to the primary sources on which an tie is based on and thereby makes it easier to understand what an tie and a node represent. Many of the documents stored in our collection come with a date of creation. This allows us to filter the network so it only displays connections of documents created within certain time spans. This timeline also indicates the number of photos per date in the collection.

Another filtering option is the number of connecting documents, which allows the visualization of those relationships that are only included in an interval of a minimum and maximum number of documents. This feature is

useful to highlight frequent co-occurrences. Finally, the number of appearances of a person in the processed collection lets us identify people who appear particularly often in any given time frame.

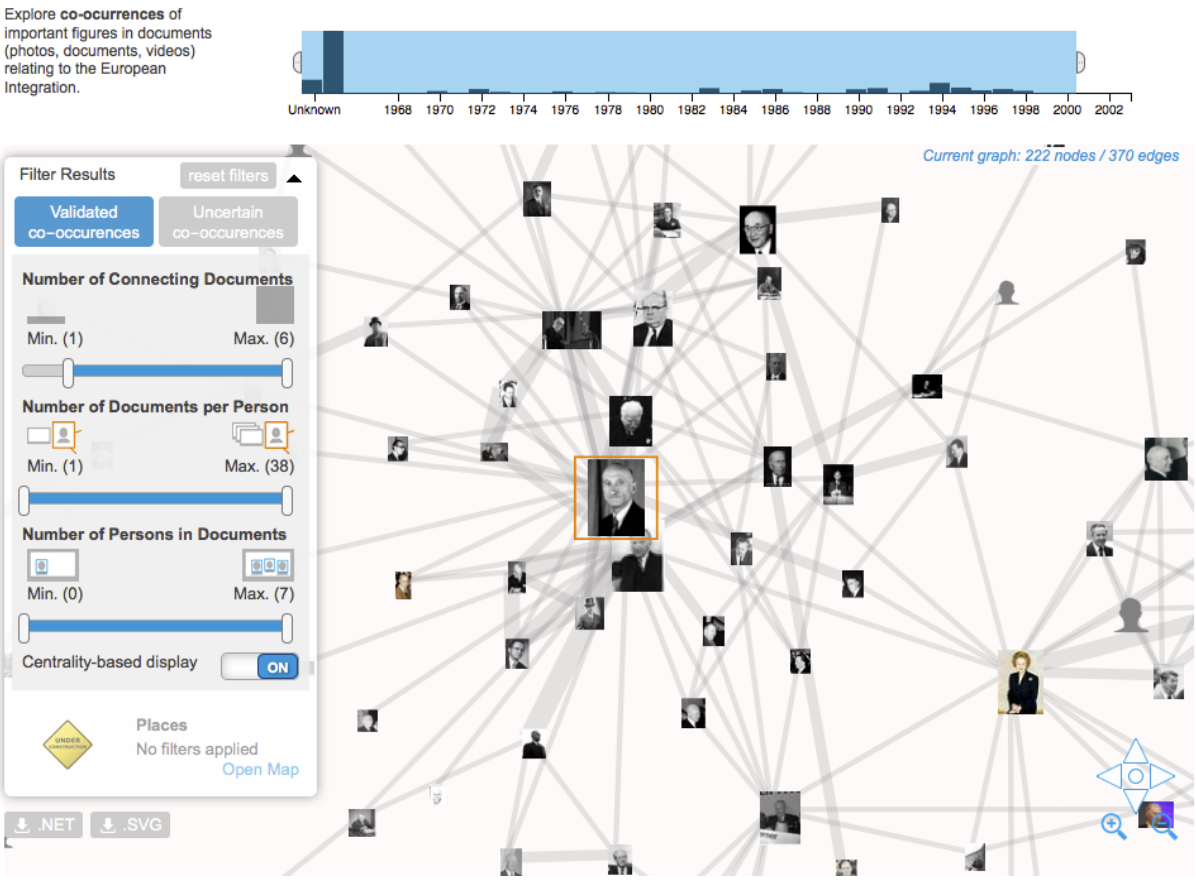

Figure 2: Screenshot of histoGraph's network visualization tool.

Another challenge for histoGraph and the Digital Humanities in general is the conception of truth. Scientists can rely on a more or less stable foundation of what is true: Any experiment can be replicated and measured precisely. In the humanities the concept of truth is far more complex: It is based on the insight, that there is no neutral or objective way to study human environments. The way in which questions are asked, when, by whom, how data is selected to answer them, by what means this data is analyzed and finally the way in which the results of such analyses are communicated and received all challenge the idea of “one truth”. Persons may, for example, change careers, their home countries might be renamed or they choose to go by different names. This means that at different points in time there might be more than one “true” answer to the simple question “Who is this?”.

Figure 3: Screenshot of histoGraph's tie contextualization.

In order to represent the discursive nature of truth in the humanities we make use of a community-driven tool for question answering, similar to stackoverflow.com (*Stack Overflow*). Users have the opportunity to answer questions and thus benefit from the knowledge within the expert crowd. However, the system allows for more than one answer and offers its users the possibility to vote and answer up or down, thereby allowing more than one answer to enter in competition with each other whilst also maintaining the full spectrum of the discussion.

2 The Semantic Web in the cultural heritage domain

Network visualizations such as those developed for histoGraph may at first sight be seen in close proximity to buzzwords such as “Semantic Web”, “Linked Data” and “Social Network Analysis”. And to some extent they want the same thing: Link cultural heritage objects with the goal to ease and improve content discovery and analysis. In order to understand the status quo and inherent problems in the use of network visualizations in cultural heritage we begin by defining and contrasting these terms.

The term “Semantic Web” was coined by Tim Berners-Lee, James Hendler and Ora Lassila (Berners-Lee, Hendler, Lassila 2001) in a paper published by the Scientific American and stands for an envisioned evolution of the World Wide Web. In the WWW, machines remain ignorant of the information contained in for example a web page. They can tell what is a heading and what is a paragraph but have no way to determine what they are about. Berners-Lee et al. propose to use ontologies to describe elements and to equip them with unique identifiers. These can be linked automatically and processed by machines based on metadata standards. In the cultural heritage sector today, this vision has been implemented to an extent in form of links between digital collections. Linked Data, another buzzword, “describes a method of publishing structured data so that it can be interlinked and become more useful through semantic queries” (*Linked data* 2015) Notable examples are Pelagios (*Pelagios*) and the recently re-launched online version of the Deutsche Biographie (*Deutsche Biographie*).

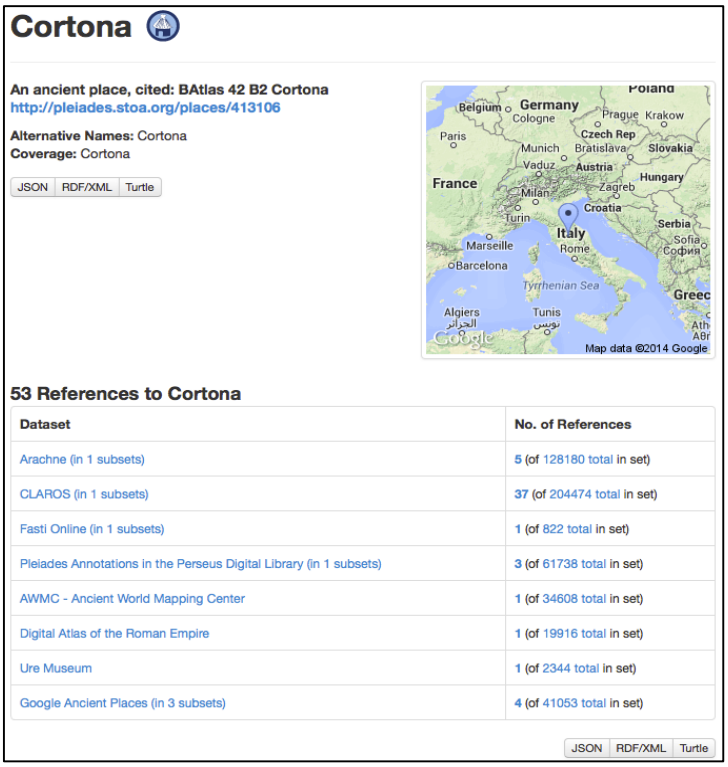

Figure 4: List of resources by geo-location in Pelagios.

In Pelagios, users select a site on a map of the Roman Empire and retrieve a list of links to associated objects kept by dozens of institutions all over the world (see Figure 1). Deutsche Biographie provides a similar service and links a person's entry to associated photos, pictures or texts provided by other institutions. Both services rely on Linked Data principles in order to save users from having to search for related information in a large number of places.

3 Social Network Analysis and Cultural Heritage

Network visualizations are powerful means to reveal complex relations between a large number of (social) entities such as for example the flow of communication between a group of co-workers. Scholars in the field regularly point out that for all their attractiveness and capabilities, network visualizations can also be suggestive and misleading for both its creator and audience, or in the words of Ben Fry: „There is a tendency when using graphs to become smitten with one's own data. (...) Graphs have a tendency of making a data set look sophisticated and important, without having solved the problem of enlightening the viewer." (Fry 2007, p. 240).

Network analysis projects are typically question-driven. Concrete research questions shape the underlying data model and the conceptualization of nodes and ties. The development of such a data model takes time, must be informed by hypotheses and in-depth knowledge of the source data.

Relations between entities can be described mathematically. Different conceptualizations of a node's influence (or *centrality*) or clusters of nodes in a network can be expressed algorithmically, see for example Wasserman, Faust 1994 for a thorough introduction. Scholars use such network visualizations and algorithms to systematically study social relations in order to explain their possible effects on human behaviour. Network visualizations often take the form of node-link-diagrams (see Figure 5), which can help understand the complex relations between any number of

entities. Other visualization techniques such as matrices (see Figure 6) can be more effective especially for large datasets (Ghoniem, Fekete, Castagliola 2004) but are less common.

The analysis of network data requires a good knowledge of the underlying dataset, its inherent biases and the circumstances of its creation. In most cases, networks do not speak for themselves but raise new questions for those who know how to read and manipulate them (Krempel 2005). There is no lack of tools for the visual and computational analysis of network data for different use cases. Their abilities range from manual network-drawing (*VennMaker*, *EgoNet*) to basic browser-based exploratory network and spatial visualization solutions (*Palladio*), to general purpose tools for data analysis and visualization (*NodeXL* [no date], *Gephi* [no date], *visone* [no date], *Pajek Wiki* [no date], *UCINET Software* [no date]) to specialized solutions for the analysis of dynamic networks RSiena (*RSiena*) to programming packages such as NetworkX (*NetworkX*), Graph-tool (*graph-tool*).

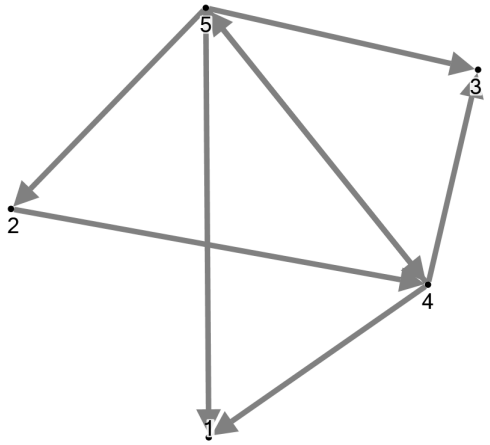

Figure 5: Example of a Node-Link-Diagramme.

	A	B	C	D	E	F
1		1	2	3	4	5
2	1	0	1	1	0	1
3	2	1	0	1	1	0
4	3	1	0	0	0	0
5	4	1	0	1	0	1
6	5	1	1	1	1	0
7						

Figure 6: Representation of the same information in form of a matrix.

Domain-specific questionnaires (for example Hogan, Carrasco, Wellman 2007; Schönhuth 2012; Schiffer, Hauck 2012 but also the abundance of born-digital data on human communication and other networks (see *Stanford Large Network Dataset Collection* for examples) make it relatively easy for social scientists to collect and process relevant data.

The generation of network data for research on subjects in the humanities and cultural heritage domain is typically more laborious. There are many different approaches to create and conceptualize network datasets, too many to reproduce an exhaustive list here. We therefore limit ourselves to illustrative examples: Eumann, März 2012 and Düring 2015 manually extracted network data from texts which yields relatively small but question-specific datasets. Stark 2012, Lemercier, Rosental 2009 and Clemens, Reupke 2008 extract networks from serial sources. Schich et al. use cultural heritage metadata, namely the places of birth and death of notable personalities in the last 2000 years, to study migration and to reconstruct the rise of cultural centres (Schich, Song, Ahn, Mirsky, Martino, Barabási, Helbing 2014), ePistolarium, a tool developed in the context of the project Circulation of

Knowledge and Learned Practices in the 17th-century Dutch Republic combines faceted search of letter correspondences with network and map-based visualizations (Huygens ING 2013) and so do two other notable projects work on Early-Modern letters: *Mapping the Republic of Letters* and *Cultures of Knowledge*. None of these projects relied on ready-to-use metadata but instead had to allocate substantial resources for the preparation of their data.

Ready-to-use metadata of historical objects which can be used to answer original network-driven research questions is hard to come by. Data availability, data quality requirements, copyright issues and the goal, to discover something new, almost mutually exclude each other. An ideal dataset should (1) contain relations between entities, which have explanatory value, (2) contain relations between entities, which are of homogeneous nature, i.e. comparable and (3) should be cohesive so as to allow the study of a network of entities, not just a number of individual cases. Most metadata do not match these criteria. They are either too trivial, incomparable, small and/or fail to capture the desired relations.

Another way to generate network data are co-occurrence networks. Co-occurrence networks are based on the following rationale: If two entities co-occur *n* times, we can reason that they are connected in some way without being able to further specify how. They are typically the result of fully or semi-automatically created networks (see Diesner, Carley, Tambayong 2012; van de Camp, van den Bosch 2012; *ConText* and Frederik Elwert's contribution in this issue). Recent research in computational linguistics tries to further distinguish between for example positive and negative relationships but has not yet reached sufficient robustness to be widely applicable. From a humanities perspective, it is hard to make non-trivial statements about a network based on co-occurrences alone. At the same time, co-occurrence can be a starting point for a human verification of automatically detected entities and a further, question-driven specification of relationships between them. A keyword-in-context display of documents and the option to download any relevant materials can support the creation of question-driven (network) data substantially.

4 Network visualizations in the cultural heritage domain

Quantitative analyses of any kind require representative data sets of their objects of study. CVCE's and the collections of many other cultural heritage institutions were not developed with this application in mind. Rather, they are closer to the hermeneutic ideal of complementing sources. This means that one source helps the interpretation of another. Any interpretation is based on this coupling of documents. Rather than treating network relations as an *explanans*, most cultural heritage institutions make use of network visualizations as a mere means to discover relations between entities.

Metadata make it relatively easy to create them and most visualizations typically allow users to interact with them. We observe three types of interactive capacities of visualizations in the cultural heritage domain today:

1) Basic: In most cases visualizations are used as a supplementary visual representation of linked information (examples are *LONSEA - League of Nations Search Engine*, *VIAF Dataset*, *People of the Founding Era*). They make it easier for users to understand relations between entities but are typically not interactive. Figure 7: Screenshot of an interactive network visualization by People of the Founding Era. shows an ego network taken from *People of the Founding Era* [no date] . Relations between people, places or institution are often represented as a so-called star network. The potential of this approach is obvious: A step-by-step discovery of related objects and the chance to discover unexpected shared entities. Albeit maybe not suitable for any mathematical description, such visualizations can have still value since they may reveal the unexpected existence or absence of relations and clusters of entities. We are not aware of any attempts to use networks to create links between objects across collections.

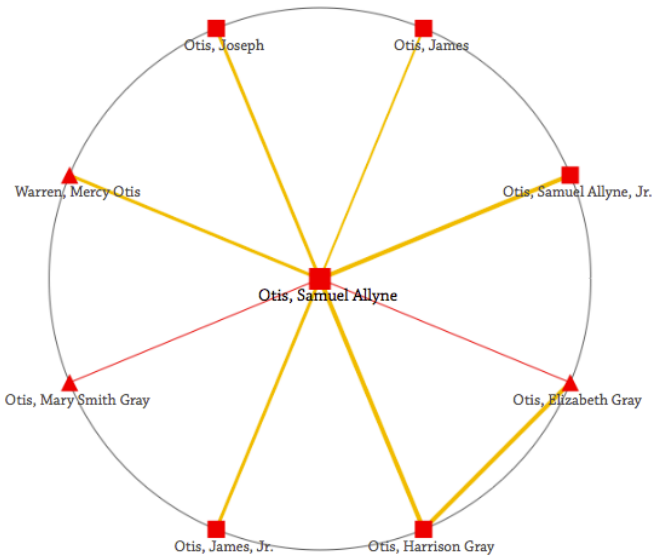

Figure 7: Screenshot of an interactive network visualization by People of the Founding Era.

2) Navigation: The visualization displays relations between items in a repository and lets users actively discover new ones (e.g. Huygens ING 2013; *SNAC*). They decide which nodes are shown in the network and are able to add additional nodes to further explore the evolving network and toggle different visualization methods such as networks or maps as shown in

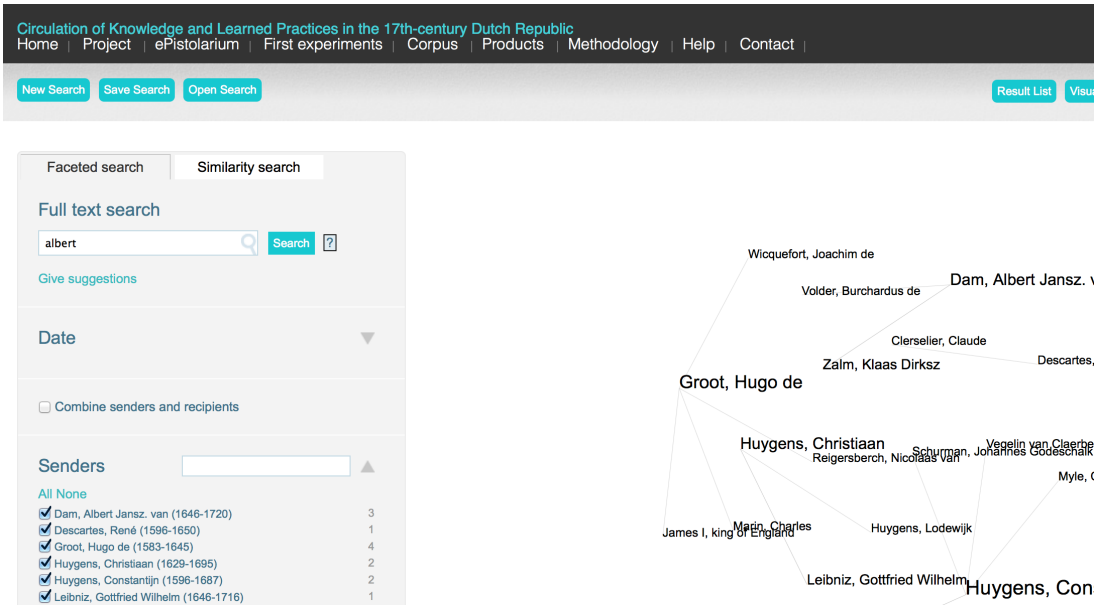

Figure 8.

Figure 8: Screenshot of ePistolarium's network drawing tool.

3) Network creation: Users are relatively free to decide what constitutes a node in a network and to define what constitutes an tie between them. A good example is the HuNI project (*Humanities Networked Infrastructure*). It provides users with the ability to define relations between objects from 30 Australian culture and cultural heritage organizations. Users can create their own collections and specify links between objects. HuNI offers a range of predefined links such as "Acted in", "Director of" or "influenced By" but also allows the definition of new types of relationships. Notes can be attached to each link to further specify the meaning of a relationship.

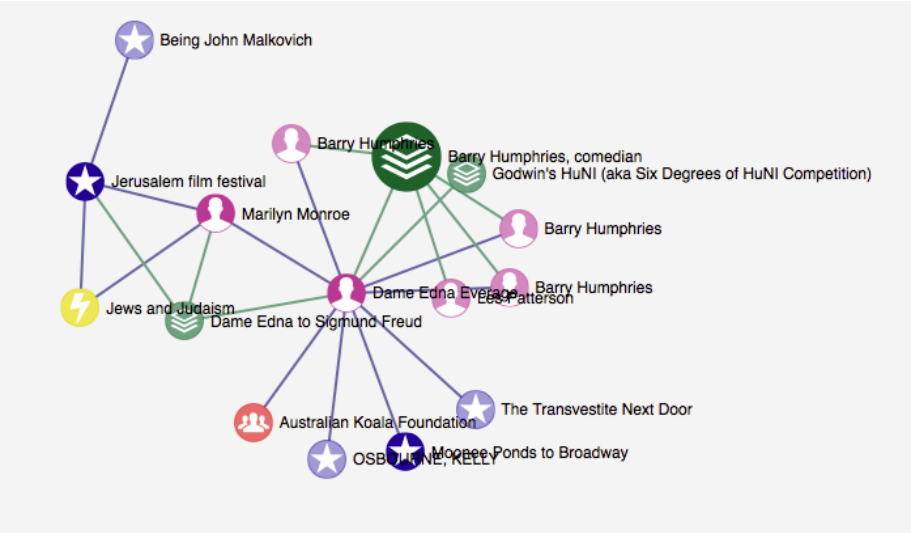

Figure 9: Screenshot of a user-generated HuNI collection showing links between persons, collections, events and concepts.

Figure 9 shows a screenshot of a user-generated collection in HuNI (including multiple occurrences of the person "Barry Humphries"). A double-click on any node expands to network and reveals that node's relations. While this does provide a comprehensive view on the relations between a small number of nodes, we suspect that the multiple types of actors and ties will result in a highly complex network which will become overwhelming after a few clicks. HuNI may still be a bit rough around the ties but it already succeeds in taking user-generated collections to the next level: Adding objects to collections is good, showing how they relate to each other is better and HuNI's vision to have user-generated content constitute an additional layer of relations between Australian cultural heritage objects is the logical consequence of this concept. A trans-institutional implementation of this vision is desirable.

5 Network visualizations in data journalism

Journalists have begun to embrace network visualizations and to an extent network computations in the increasingly popular field of data journalism. They share many of the challenges cultural heritage institutions face: Limited technological skills and interest on the side of creators and audiences and the urgent need for comprehensible, interactive, engaging visual representations. In this section we will (1) present a small number of best practices in the integration of network visualizations in journalistic publications (based on this compilation: *Untangled: Investigations*)) and (2) discuss tools which have been developed for journalists inasmuch as they are relevant for the cultural heritage domain.

In a special edition on the 2010 US American elections, NPR published a flow-chart inspired network visualization which revealed relations between organizations associated with the Republican party despite the fact that "[t]hese groups style themselves as independent, grass-roots organizations"(*A Web Of GOP Influence, Network Diagrams Are Hard - Features - Source: An OpenNews project*)

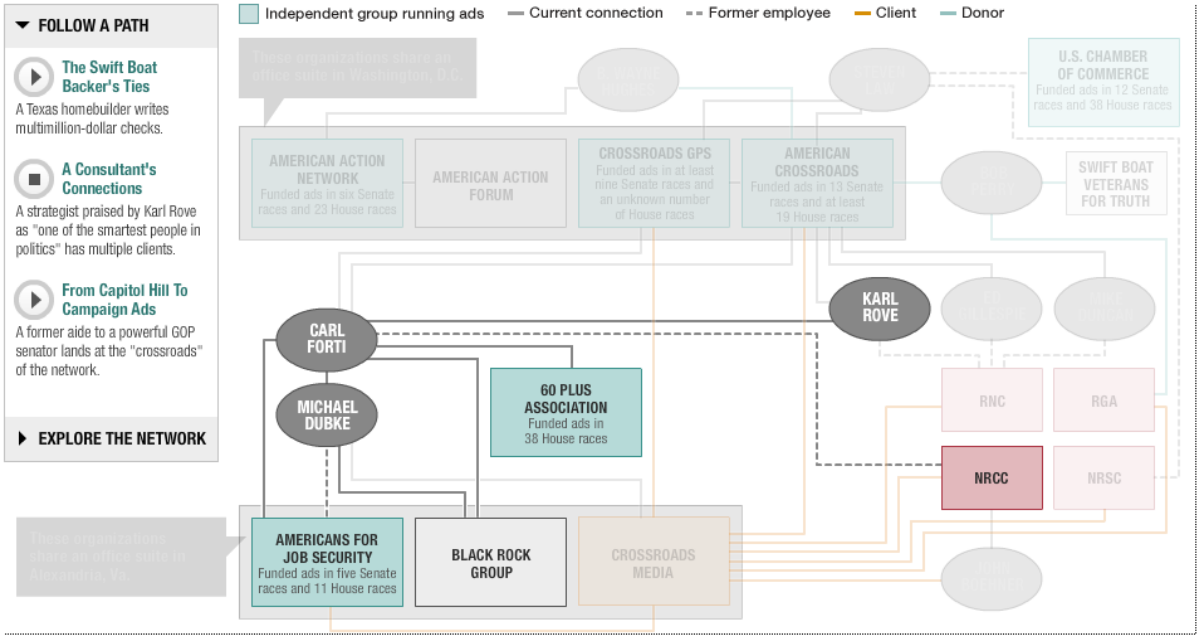

Figure 10: Screenshot of NPR's "A Web of GOP Influence" article.

A rare feature in most similar visualizations, the NPR provides three tours through the network. During a tour a narrator describes the significance of relationships between entities which are highlighted in the course of the story. This guidance provides much needed orientation and adds meaning to the complex set of nodes and ties. It thereby communicates the meaning behind the visualization whilst still providing the opportunity for users to explore the network by themselves.

The Los Angeles Times published an article title "The Calderon family's connections" (Lutz) which links network visualizations and accompanying visualizations in two ways. A click on the black boxes below the title (see Figure 11) reveals how the family members relate to the respective organization but also leads the reader to the corresponding section of the article. Each section provides facts about the actors involved and how they relate to each other. This is accompanied by a visual representation of these relations on the left side which is automatically updated when users scroll to a different section of the article. A click on a node in this network provides context information.

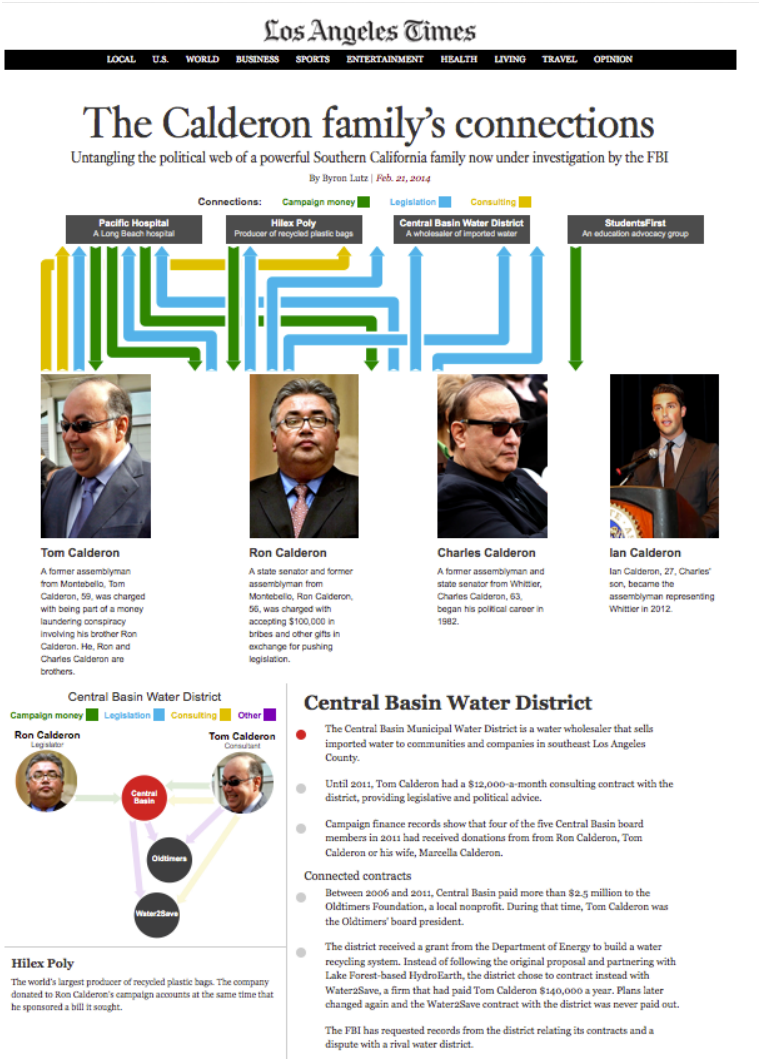

Figure 11: Screenshot of the article "The Calderon family's connections.

The platform LittleSis developed Oligrapher (*Oligrapher*), a tool which displays multiple types of relations between a small number of persons and institutions based on crowd-sourced data which provide orientation but require accompanying contextualization.

A number of tools seek to support journalists with the creation and visualization of networks. Again, we focus on a small collection and highlight features which are of relevance for our domain. Detective.io (*Detective.io*) is a tool which highlights relations between any dataset provided by the user in form of a csv file and based on their definitions of nodes and ties. The tool makes it very easy to select different types of actors and to reveal the relations between them. A click on a node reveals attribute information and its ego network (see Figure 12).

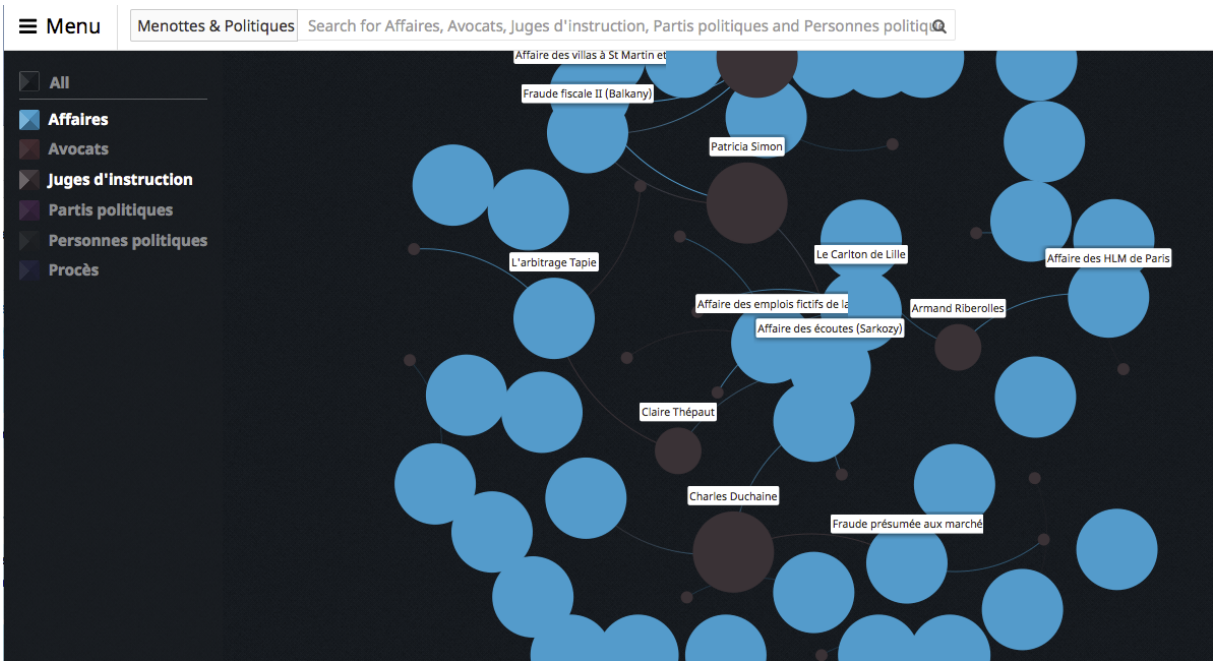

Figure 12: Screenshot taken from Detective.io.

Graph Commons (*Graph Commons*) supports the crowd-based collection of relational data and tracks individual contributions. Kumu (*Kumu*) lets users create, filter and cluster network visualizations and supports the annotation of ties between nodes. Médialab's Hyphe tool (Hyphe) is still in development but promises to provide a service for the semi-automatic creation of network visualization based on web contents. Thomson Reuter's Connected China (ConnectedPRC) provides timelines, network inspired visualizations and interactive maps with background information and articles on China's government, development and politicians (see Figure 13). It is a resource very rich in information which displays multiple types of ties between multiple types of nodes but requires users to learn how to use it and the navigation can be confusing at times.

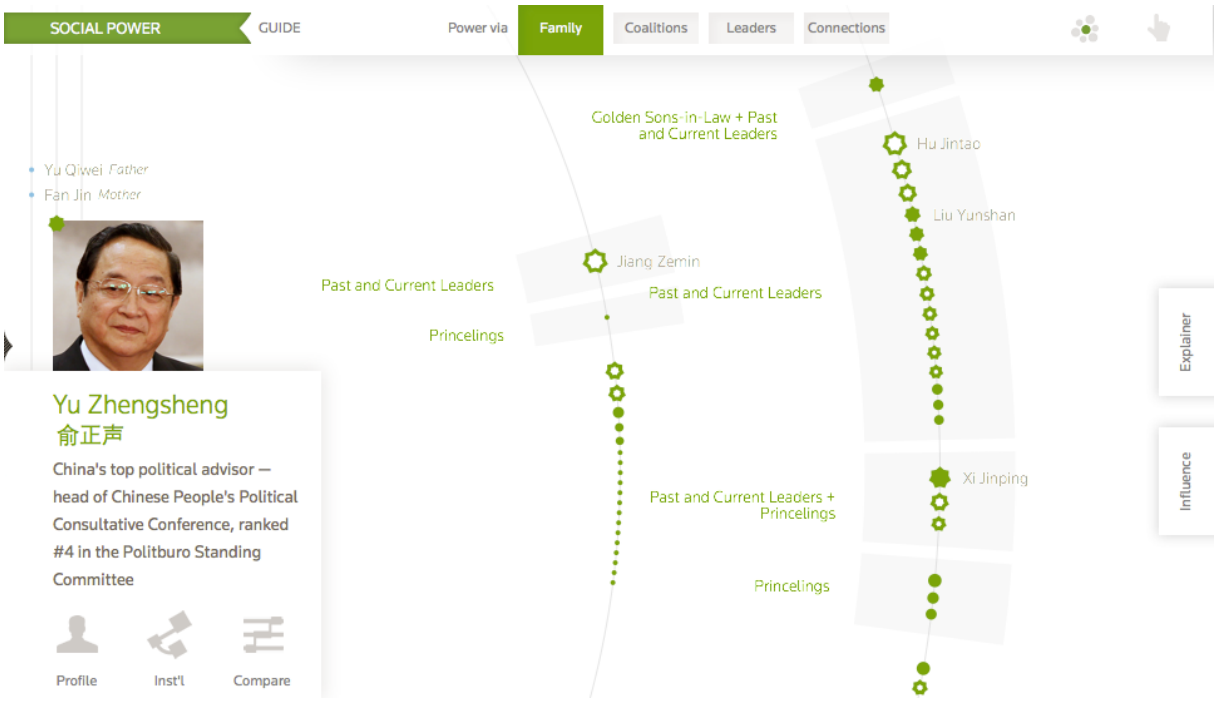

Figure 13: Screenshot taken from Connected China.

6 Application at CVCE

Our survey reveals that (1) Recent advances in the realisation of the Semantic Web vision in cultural heritage data does translate into opportunities for network analysis projects. Such projects still require concrete research interests and a high degree of commitment. (2) There is no lack of tools for the computational and visual analysis of network data. (3) Existing network visualizations in the cultural heritage domain provide basic functionalities but quickly leave users disoriented. (4) Applications developed in the data journalism domain simplify the process of creating annotated network visualizations and some of them successfully combine them with other media such as text or images.

Network visualizations can be treacherous and require that viewers understand their intended purpose, the conceptualization of the data and possible biases. Even more important: viewers need to understand why they should look at one and what can discover. Most network visualizations are still published in the tradition of the print-era: As static images which are explained by the surrounding text and a legend. Network animations which display change over time, highlight areas of interest and provide additional contextual information offer much needed orientation, and help communicate the epistemic value of a visualization. To our knowledge there is no tool which can bridge the gap between network visualization and storytelling in the way that for example StorymapJS (*StoryMap JS*) and Story Map (*Story Maps*) do this for spatial data.

We envision that histoGraph fills this niche and becomes a general purpose context exploration and storytelling tool for research and teaching in cultural heritage, the humanities and journalism. Crucially, histoGraph will maintain its current ability to process photos and will also become capable of processing text documents and metadata. We seek to integrate it in an existing tool, CVCE MyPublications. MyPublications allows registered users to create their own collections of documents. At this stage, these can be viewed as slideshows (see *MyEurope* [no date] for a similar solution).

With regard to the future development of histoGraph we propose the following use cases: Network creation based on images, network creation based on entity co-occurrences, network annotation and storytelling. This will require the following components:

1) Redevelopment of the image processing components with open source solutions and enhanced usability
2) Development of a component which imports collections stored in CVCE's MyPublication tool and uses language processing technology to automatically produce visualizations of co-occurring entities such as persons, institutions, places.
3) Development of browser-like "Back" and "Forward" buttons, a dynamically updated node list as well as logs of reversible user actions as implemented in Open Refine (*OpenRefine*) for better orientation.
4) Development of an annotation component which allows users to manually link nodes and ties to (third party) source documents and to provide additional textual information
5) Redevelopment of the contextualization feature which visualizes such links to source documents for any node and tie (see XXXX)
6) Development of a "guided tour" component which lets users create storylines based on annotations in the network and accompanying text. This will utilize the network animations outlined above
7) Development of an import functionality for network data (e.g. social, word, metadata-based or technical networks) in the gexf and graphml file formats which recognizes attributes (source documents, node size/color, tie width, clustering etc.) generated by other network analysis tools and processes and allows their annotation as outlined in (3) and (5)
8) Development of map-based visualizations of data and hybrid map+network visualizations.

These are ambitious goals, which will require time, funding and a number of case studies for the development of prototypes. Network visualizations have unrivalled powers to make complex relations comprehensible – we hope that histoGraph will allow authors and audiences to make the most of this potential.

7 Summary

In this paper we presented histoGraph, a tool for the extraction of networks from historical photographs, surveyed the current application of social network analysis and network visualizations in the cultural heritage sector and data journalism. We find that there is a gap between the visualization and analysis of network data on one side and the need for contextualization of such data and storytelling practices on the other. We therefore outline milestones, which will help us to bridge this gap.

References

A Web Of GOP Influence, [no date]. NPR.org [online], [Accessed 15 January 2015]. Available from: http://www.npr.org/2010/10/27/130845545/a-web-of-gop-influence

BERNERS-LEE, Tim, HENDLER, James and LASSILA, Ora, 2001, The Semantic Web. Scientific American. 2001. Vol. 284, no. 5, p. 34–43.

Collaborative “network mapping” platform and knowledge base of relationships, [no date]. Graph Commons [online], [Accessed 15 January 2015]. Available from: http://graphcommons.com

ConText, [no date]. [online], [Accessed 13 January 2015]. Available from: http://context.lis.illinois.edu/

CUbRIK Project, 2014. [online], [Accessed 12 January 2015]. Available from: http://www.cubrikproject.eu/

Cultures of Knowledge | Networking the Republic of Letters, 1550-1750, [no date]. [online], [Accessed 12 January 2015]. Available from: http://www.culturesofknowledge.org/

Detective.io, [no date]. [online], [Accessed 16 January 2015]. Available from: https://www.detective.io/tour/

Deutsche Biographie, [no date]. [online], [Accessed 12 January 2015]. Available from: http://www.deutsche-biographie.de/

DIESNER, Jana, CARLEY, Kathleen M. and TAMBAYONG, Laurent, 2012, Extracting socio-cultural networks of the Sudan from open-source, large-scale text data. Computational and Mathematical Organization Theory [online]. September 2012. Vol. 18, no. 3, p. 328–339. [Accessed 2 August 2014]. DOI 10.1007/s10588-012-9126-x. Available from: http://link.springer.com/10.1007/s10588-012-9126-x

DÜRING, Marten, 2015, Verdeckte soziale Netzwerke im Nationalsozialismus. Berliner Hilfsnetzwerken für verfolgte Juden. Berlin : De Gruyter.

EgoNet, [no date]. [online], [Accessed 16 January 2015]. Available from: http://socioworks.com/productsall/egonet/

EUMANN, Ulrich and MÄRZ, Jascha, 2012, Das Schneeballsystem der Gestapo im Kampf gegen den Widerstand. Österreichische Zeitschrift für Geschichtswissenschaft. 2012. Vol. 23, p. 323–340.

FRY, Ben, 2007, Visualizing Data: Exploring and Explaining Data with the Processing Environment. 1 edition. Beijing; Cambridge : O’Reilly Media. ISBN 9780596514556.

Gephi, [no date]. [online], [Accessed 16 January 2015]. Available from: https://gephi.github.io/

GHONIEM, M., FEKETE, J. and CASTAGLIOLA, P., 2004, A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations. Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on. 2004. P. 17–24. DOI 10.1109/INFVIS.2004.1.

Graph Commons, [no date]. [online], [Accessed 16 January 2015]. Available from: http://graphcommons.com/

graph-tool, [no date]. [online], [Accessed 16 January 2015]. Available from: http://graph-tool.skewed.de/

HistoGraph, [no date]. [online], [Accessed 12 January 2015]. Available from: http://www.histograph.eu

HOGAN, Bernie, CARRASCO, Juan Antonio and WELLMAN, Barry, 2007, Visualizing Personal Networks: Working with Participant-aided Sociograms. Field Methods [online]. 1 May 2007. Vol. 19, no. 2, p. 116–144. [Accessed 13 January 2015]. DOI 10.1177/1525822X06298589. Available from: http://fmx.sagepub.com/content/19/2/116

Humanities Networked Infrastructure - HuNI, [no date]. [online], [Accessed 13 January 2015]. Available from: https://huni.net.au/#/search

HUYGENS ING, 2013, ePistolarium. ePistolarium [online]. 10 March 2013. [Accessed 3 October 2013]. Available from: http://ckcc.huygens.knaw.nl/epistolarium/

KREMPEL, Lothar, 2005, Visualisierung komplexer Strukturen. Frankfurt am Main : Campus Verlag GmbH. ISBN 9783593378138.

Kumu, [no date]. [online], [Accessed 16 January 2015]. Available from: https://kumu.io/

Linked data, 2015. Wikipedia, the free encyclopedia [online], [Accessed 8 January 2015]. Available from: http://en.wikipedia.org/w/index.php?title=Linked_data&oldid=641307774
Page Version ID: 641307774

LONSEA - League of Nations Search Engine, [no date]. [online], [Accessed 26 September 2013]. Available from: http://www.lonsea.de/

LUTZ, Byron, [no date], The Calderon Family's Connections. graphics.latimes.com [online]. [Accessed 15 January 2015]. Available from: http://graphics.latimes.com/calderon/

Mapping the Republic of Letters, [no date]. [online], [Accessed 12 January 2015]. Available from: http://republicofletters.stanford.edu/index.html

MyEurope, [no date]. [online], [Accessed 13 January 2015]. Available from: http://www.inventingeurope.eu/myeurope/

Network Diagrams Are Hard - Features - Source: An OpenNews project, [no date]. [online], [Accessed 15 January 2015]. Available from: https://source.opennews.org/en-US/articles/network-diagrams-hard/

NetworkX, [no date]. [online], [Accessed 15 February 2014]. Available from: http://networkx.github.io/

NodeXL, [no date]. CodePlex [online], [Accessed 16 January 2015]. Available from: http://nodexl.codeplex.com/Wikipage?ProjectName=nodexl

Oligrapher, [no date]. [online], [Accessed 14 January 2015]. Available from: http://blog.littlesis.org/2014/07/22/introducing-oligrapher-power-mapping-on-littlesis/

OpenRefine, [no date]. [online], [Accessed 13 January 2015]. Available from: http://openrefine.org/

Pajek Wiki, [no date]. [online], [Accessed 15 February 2014]. Available from: http://pajek.imfm.si/doku.php

Palladio, [no date]. [online], [Accessed 19 August 2014]. Available from: http://palladio.designhumanities.org/#/

Pelagios, [no date]. [online], [Accessed 8 January 2015]. Available from: http://pelagios-project.blogspot.co.uk/

People of the Founding Era, [no date]. [online], [Accessed 12 January 2015]. Available from: http://pfe.rotunda.upress.virginia.edu/

RSiena, [no date]. [online], [Accessed 16 January 2015]. Available from: http://www.stats.ox.ac.uk/~snijders/siena/siena_r.htm

SCHICH, Maximilian, SONG, Chaoming, AHN, Yong-Yeol, MIRSKY, Alexander, MARTINO, Mauro, BARABÁSI, Albert-László and HELBING, Dirk, 2014, A network framework of cultural history. science. 2014. Vol. 345, no. 6196, p. 558–562.

SCHIFFER, Eva and HAUCK, Jennifer, 2012, Between intuition and indicators – Using Net-Map for visual and qualitative social network analysis. In : Knoten und Kanten 2.0: soziale Netzwerkanalyse in Medienforschung und Kulturanthropologie. p. 231 – 257. ISBN 9783837619270 3837619273.

SCHÖNHUTH, Michael (ed.), 2012, VennMaker. Wiesbaden : VS Verlag für Sozialwissenschaften.

SNAC, [no date]. [online], [Accessed 4 September 2013]. Available from: http://socialarchive.iath.virginia.edu/

Stack Overflow, [no date]. [online], [Accessed 12 January 2015]. Available from: http://stackoverflow.com/

Stanford Large Network Dataset Collection, [no date]. [online], [Accessed 13 January 2015]. Available from: http://snap.stanford.edu/data/

STARK, Martin, 2012, Soziale Einbettung eines ländlichen Kreditmarktes im 19. Jahrhundert [online]. [Accessed 16 January 2015]. Available from: http://ubt.opus.hbz-nrw.de/volltexte/2014/839/

StoryMap JS, [no date]. [online], [Accessed 16 January 2015]. Available from: http://storymap.knightlab.com/

Story Maps, [no date]. [online], [Accessed 16 January 2015]. Available from: http://storymaps.arcgis.com/en/

UCINET Software, [no date]. [online], [Accessed 15 February 2014]. Available from: https://sites.google.com/site/ucinetsoftware/home

Untangled: Investigations, [no date]. [online], [Accessed 14 January 2015]. Available from: http://untangled.knightlab.com/investigations.html

VAN DE CAMP, Matje and VAN DEN BOSCH, Antal, 2012, The socialist network. Decision Support Systems. 2012. Vol. 53, no. 4, p. 761–769. DOI 10.1016/j.dss.2012.05.031.

VennMaker, [no date]. [online], [Accessed 16 January 2015]. Available from: http://www.vennmaker.com/

VIAF Dataset, [no date]. [online], [Accessed 10 May 2012]. Available from: http://viaf.org/viaf/data/

visone, [no date]. [online], [Accessed 15 February 2014]. Available from: http://visone.info/

WASSERMAN, Stanley and FAUST, Katherine, 1994, Social Network Analysis: Methods and Applications. New York : Cambridge University Press. Structural Analysis in the Social Sciences. ISBN 978-0521382694.

WIENEKE, Lars, DÜRING, Marten, SILAUME, Ghislain, LALLEMAND, Carine, CROCE, Vincenzo, LAZZARRO, Marilena, NUCCI, Francesco, PASINI, Chiara, FRATERNALI, Piero, TAGLIASACCHI, Marco, MELENHORST, Mark, NOVAK, Jasminko, MICHEEL, Isabel, HARLOFF, Erik and GARCIA MORON, Javier, 2014, histoGraph – A Visualization Tool for Collaborative Analysis of Historical Social Networks from Multimedia Collections. In : Proceedings of 18th International Conference Information Visualisation (IV), 2014 Conference. Paris, France.

Biography of the authors

Marten Düring works as a researcher in the Digital Humanities Lab at Centre Virtuel de la Connaissance sur l'Europe (CVCE), Luxembourg. At CVCE he is responsible for the network analysis research strand and the redevelopment of histoGraph, a tool for the crowd-based annotation of multimedia sources and their graph-based representation. A historian by training (BA Cultural History 2004 in Augsburg, MA War, History, Memory 2006 in Manchester, PhD Contemporary History 2012 in Mainz), his work has previously focused on helping behaviour during the Holocaust and memories of aerial warfare during the Second World War. Between 2012/2013 he worked as eHumanities guest Post Doc with Antal van den Bosch at Radboud University Nijmegen. Between January and June 2014 he held the position of CDHI Digital History Postdoc at UNC Chapel Hill. Marten has published widely on helping behaviour, network analysis in the historical disciplines, the applicability of natural language processing in historical research and on the crowd-based annotation of multimedia sources.

Lars Wieneke holds a PhD in Engineering from the Bauhaus-University Weimar, Germany. After graduating from the Technical University of Ilmenau he worked as a researcher in the Department of Interface Design at the Bauhaus-University Weimar. Funded by a Marie-Curie fellowship, Lars joined the EPOCH/CHIRON research group at the University of Brighton in January 2007 and specialised in the application of user-created content for museums and cultural heritage. In 2009, Lars worked as an independent consultant for museums and cultural heritage in France before becoming project manager of a joint research project between the Jewish Museum Berlin and the University of Applied Sciences Berlin. He joined the Centre Virtuel de la Connaissance sur l'Europe in 2011 as a researcher in the Digital Humanities Lab and became the head of the information and technology department in 2014. Lars has been responsible as a work-package leader for the integration of the histoGraph vertical app in the now completed FP7 funded ICT research project CUbRIK.

Acknowledgement

Our initial work on histoGraph has been kindly supported by the FP7 funded CUbRIK project (Grant agreement no: 287704).